OPTIMAL BOUNDS IN NON-GAUSSIAN LIMIT THEOREMS FOR U-STATISTICS

V. Bentkus[1] F. Götze[1]

University of Bielefeld

June 30, 1997

Abstract. Let $X, X_1, X_2, \dots$ be i.i.d. random variables taking values in a measurable space $\mathcal{X}$. Let $\phi(x,y)$ and $\phi_1(x)$ denote measurable functions of the arguments $x, y \in \mathcal{X}$. Assuming that the kernel ϕ is symmetric and that $\mathbf{E}\,\phi(x,X) = 0$, for all x, and $\mathbf{E}\,\phi_1(X) = 0$, we consider U-statistics of type

$$T = N^{-1} \sum_{1 \le j < k \le N} \phi(X_j, X_k) + N^{-1/2} \sum_{1 \le j \le N} \phi_1(X_j).$$

It is known that the conditions $\mathbf{E}\,\phi^2(X, X_1) < \infty$ and $\mathbf{E}\,\phi_1^2(X) < \infty$ imply that the distribution function of T, say F, has a limit, say F_0, which can be described in terms of the eigenvalues of the Hilbert–Schmidt operator associated with the kernel $\phi(x,y)$. Under optimal moment conditions, we prove that

$$\Delta_N = \sup_x \bigl| F(x) - F_0(x) - F_1(x) \bigr| = \mathcal{O}\bigl(N^{-1}\bigr),$$

provided that at least 9 eigenvalues of the operator do not vanish. Here F_1 denotes an Edgeworth type correction. We provide explicit bounds for Δ_N and for the concentration functions of statistics of type T.

1. Introduction

Let $X, \overline{X}, X_1, \dots, X_N$ be independent identically distributed (i.i.d.) random variables taking values in an arbitrary measurable space $(\mathcal{X}, \mathcal{B})$. Let $\phi_1 : \mathcal{X} \to \mathbb{R}$

1991 *Mathematics Subject Classification.* Primary 62E20; secondary 60F05.

Key words and phrases. U-statistics, degenerate U-statistics, von Mises statistics, symmetric statistics, Central Limit Theorem, convergence rates, Berry–Esseen bounds, Edgeworth expansions, second order efficiency.

[1]Research supported by the SFB 343 in Bielefeld.

and $\phi : \mathcal{X}^2 \to \mathbb{R}$ be real-valued measurable functions. Assume that ϕ is symmetric, that is, $\phi(x, y) = \phi(y, x)$, for all $x, y \in \mathcal{X}$. Consider the U-statistic

$$T = \frac{1}{N} \sum_{1 \le i < j \le N} \phi(X_i, X_j) + \frac{1}{\sqrt{N}} \sum_{1 \le i \le N} \phi_1(X_i), \tag{1.1}$$

assuming that

$$\mathbf{E}\, \phi_1(X) = 0, \qquad \mathbf{E}\, \phi(x, X) = 0, \qquad \text{for all } x \in \mathcal{X}. \tag{1.2}$$

Our aim is to establish an Edgeworth expansion with remainder $\mathcal{O}(N^{-1})$, for the distribution function of T. We shall show that moment conditions on ϕ_1 and ϕ together with the assumption that the limiting weighted χ^2 distribution has at least nine degrees of freedom, are sufficient for the result to hold. We shall establish as well explicit bounds of order $\mathcal{O}(N^{-1})$ for concentration functions of a somewhat larger class of statistics than (1.1).

We shall write

$$\begin{aligned} \beta_s &= \mathbf{E}\, \bigl|\, \phi_1(X) \,\bigr|^s, & \gamma_s &= \mathbf{E}\, \bigl|\, \phi(X, \overline{X}) \,\bigr|^s, \\ \sigma^2 &= \gamma_2, & \gamma_{s,r} &= \mathbf{E}\, \Bigl(\mathbf{E} \Bigl\{ \bigl|\, \phi(X, \overline{X}) \bigr|^s \,\Bigl|\, X \Bigr\} \Bigr)^r, \end{aligned} \tag{1.3}$$

and assume throughout that

$$\beta_2 < \infty, \qquad 0 < \sigma^2 < \infty. \tag{1.4}$$

The variance of T splits as

$$\mathbf{E}\, T^2 = \beta_2 + \frac{N-1}{2\,N}\, \sigma^2. \tag{1.5}$$

The statistic T is called degenerate since the condition $\sigma^2 > 0$ ensures that the quadratic part of the statistic is not asymptotically negligible and therefore T is not asymptotically normal. More precisely, the asymptotic distribution of T is non-Gaussian and is given by the distribution of the random variable

$$T_0 = \frac{1}{2} \sum_{j \ge 1} q_j\, (\eta_j^2 - 1) + \sum_{j \ge 0} a_j\, \eta_j. \tag{1.6}$$

Here η_j is a sequence of i.i.d. standard normal random variables, $a_0, a_1, \ldots$ denotes a (non-random) sequence of square summable weights, and $q_1, q_2, \ldots$ denote eigenvalues of the Hilbert–Schmidt operator, say $\mathbb{Q}$, associated with the kernel ϕ (see Section 2 for detailed definitions). Without loss of generality we shall assume throughout that $|q_1| \ge |q_2| \ge \ldots$

Introduce the distribution functions

$$F(x) = \mathbf{P}\, \bigl\{ T \le x \bigr\}, \qquad F_0(x) = \mathbf{P}\, \bigl\{ T_0 \le x \bigr\}. \tag{1.7}$$